Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Locating the anomalous scatterer substructures in halide and sulfur phasing

Isabel Usón,[a]* Bernhard Schmidt,[b] Rixa von Bülow,[a] Susanne Grimme,[b] Kurt von Figura,[b] Miroslawa Dauter,[c] Kanagalaghatta R. Rajashankar,[c] Zbigniew Dauter[c] and George M. Sheldrick[a]

[a]Lehrstuhl für Strukturchemie, Institut für Anorganische Chemie, Universität Göttingen, Tammannstrasse 4, 37077 Göttingen, Germany, [b]Zentrum Biochemie und Molekulare Zellbiologie, Abteilung Biochemie II, Universität Göttingen, Heinrich Düker Weg 12, 37073 Göttingen, Germany, and [c]Synchrotron Radiation Research Section, National Cancer Institute, Brookhaven National Laboratory, Building 725A-X9, Upton NY 11973, USA

Correspondence e-mail: uson@shelx.uni-ac.gwdg.de

Received 7 September 2002
Accepted 14 October 2002

PDB References: Man6P/IGFII-receptor repeat 11, 1e6f, r1e6fsf; 1gqb, r1gqbsf.

Improved data quality now makes it feasible to exploit the weak anomalous signal derived only from the sulfurs inherent to the protein or in particular from halide ions incorporated by soaking. The latter technique requires the location of a high number of partially occupied halide sites. This number appears to be roughly proportional to the exposed protein surface. This paper explores the application of dual-space *ab initio* methods as implemented in the program *SHELXD* to the location of substructures of sulfur in SAD experiments, bromide in SAD and MAD experiments and iodide using SAD and SIRAS to determine the anomalous-atom substructure. Sets of atoms consistent with the Patterson function were generated as a starting point for the dual-space recycling procedure in *SHELXD*. The substructure is then expanded to the full structure by maximum-likelihood phasing with *SHARP* and density modification with the program *DM*. Success in the location of the substructures and subsequent phasing depends critically on the quality of the data and on the extent of the anomalous signal. This varies with each crystal and soak, but for the same crystal the significance of the anomalous signal was found to be highly sensitive to the redundancy of the intensity measurements, which in some cases made all the difference. This is illustrated by the determination of the previously unknown structure of repeat 11 of the human mannose-6-phosphate/insulin-like growth factor II receptor (Man6P/IGFII-receptor), with 310 amino acids in the asymmetric unit, which was phased by soaking the crystals in a cryoprotectant solution containing halide anions.

1. Introduction

The ready availability of tunable-wavelength synchrotron beamlines, together with the development of experimental techniques leading to more accurate data [cryocrystallography (Garman & Schneider, 1997), improvement in detector sensitivity, faster data acquisition allowing higher redundancy to be collected] has in recent years made phasing methods based only on the presence of anomalous scatterers very competitive with more established methods (Wang, 1985; Hendrickson, 1991).

Traditional methods of incorporating anomalous scatterers into the crystal such as substituting the methionines in the structure by selenomethionine (Smith, 1998) or soaking with metal salts and complexes work well, but the amount of work and time involved implies that a simpler, generally applicable and effective experimental way of introducing sufficient anomalous signal into the data to solve the structure would be welcome. Ideally, exploiting the anomalous signal already

Table 1
Statistics of the X-ray data collection and heavy-atom sites.

Values in parentheses are for the outermost 0.1 Å resolution shell.

Data set	Native	Native anomalous 1	Native anomalous 2	*C*2 Br^- peak	*C*2 Br^- inflection	Br^- peak	Br^- inflection	Br^- remote 1	Br^- remote 2	I^- soak
Resolution (Å)	1.75	2.30	2.38	1.79	1.79	2.50	2.50	2.10	2.50	2.20
R_{sym}† (%)	2.6 (8.9)	6.3 (49.2)	3.8 (3.14)	8.6 (34.0)	9.6 (45.2)	7.7 (37.9)	6.4 (30.2)	9.4 (66.1)	7.3 (26.0)	6.3 (24.8)
Completeness (%)	99.1 (97.4)	99.1 (91.2)	99.5 (97.0)	98.8 (95.3)	94.5 (89.8)	97.8 (78.6)	98.7 (88.9)	99.9 (99.9)	99.9 (100.0)	99.8 (98.9)
$I/\sigma(I)$	40.8	33.0 (2.9)	32.7 (8.8)	8.9 (3.0)	8.4 (2.3)	17.4 (3.3)	19.2 (3.9)	19.4 (2.7)	24.1 (7.1)	32.4 (7.5)
Redundancy	8.3	8.6	3.9	3.8	3.8	3.6	3.6	6.2	6.2	6.5

† $R_{sym} = \sum |I_{hkl} - \langle I_{hkl}\rangle| / \sum |I_{hkl}|$.

present in the native protein or in the solvent would eliminate the extra experimental work in derivatization or in the expression of selenomethionine proteins and would also eliminate the risk of lack of isomorphism.

Phasing using the anomalous signal of sulfur alone was first achieved in the case of crambin (Hendrickson & Teeter, 1981), while neurophysin II was phased using only the anomalous signal derived from a monoiodinated peptide bound in the hormone-binding site (Chen *et al.*, 1991). The structure of tetragonal lysozyme could also be phased from the combined anomalous signal of the ten sulfurs in the protein and seven coordinated chloride anions originating from the crystallization buffer (0.1 *M* NaCl) with highly redundant (mean redundancy of 22) synchrotron data collected at a wavelength of 1.54 Å (Dauter *et al.*, 1999).

The contribution of the halide anions in the case of lysozyme can be further exploited, as they can be substituted by their higher homologues bromine and iodine either by replacing the NaCl in the crystallization buffer by NaBr (Dauter & Dauter, 1999) or by a short soak in a cryobuffer containing the halide salts (Dauter *et al.*, 2000).

One limitation in this method is the need to locate a large number of partially occupied halide sites; this number appears to be proportional to the exposed protein surface, as binding is neither quantitative nor specific with respect to the coordination geometry and ligands. On the other hand the binding is chemically fairly specific, with the same sites around two monomers in the asymmetric unit (Devedjiev *et al.*, 2000; Dauter *et al.*, 2000). Therefore, the complexity of this step is expected to increase with the number of amino acids in the asymmetric unit, in many cases exceeding the capacity of conventional Patterson and direct methods.

Dual-space recycling methods (Miller *et al.*, 1993; Usón & Sheldrick, 1999), which were originally developed for *ab initio* phasing of macromolecules at high resolution, have proven useful in the determination of heavy-atom or anomalous scatterer substructures at medium (∼3 Å) resolution (Sheldrick & Schneider, 2001; Howell *et al.*, 2000). They have the advantage over conventional Patterson or direct methods of being more effective in solving larger structures or substructures and being less susceptible to missing or aberrant reflections. They also provide a way of integrating direct and Patterson methods.

This paper uses the previously unknown protein structure of repeat 11 (the IGFII binding repeat) in the Man6P/IGFII receptor (subsequently referred to as R11), with 310 amino acids in the asymmetric unit, to explore the application of dual-space *ab initio* methods to the location of substructures of anomalous sulfur, bromine in single-wavelength anomalous (SAD) and MAD experiments and anomalous iodine (SAD and SIRAS) to find a starting substructure leading to phasing of the final structure when coupled with maximum-likelihood phasing with *SHARP* (de La Fortelle & Bricogne, 1997) and density modification with *DM* (Cowtan, 1994).

After this work had been completed and the coordinates deposited with the PDB (codes 1e6f and 1gqb), the structures of two different crystal forms, each with a monomer in the asymmetric unit as opposed to the dimers found in the two forms reported here, were published (Brown *et al.*, 2002).

2. Materials and methods

2.1. Production and crystallization

R11 was expressed and purified as described elsewhere (Grimme *et al.*, 2000). Single crystals were grown by vapour diffusion at 288 K in hanging drops (10 µl) from a protein solution containing 8 mg ml^{-1} protein in 10 m*M* Tris–HCl buffer pH 7.5 and 150 m*M* NaCl, using 28% PEG 4000 and 0.2 *M* ammonium acetate in 100 m*M* cacodylate buffer pH 5.6–5.8 as a precipitant in a 1:1 ratio. The crystals grew over a period of one month. Under these conditions, two kinds of crystals were obtained. One form belongs to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 48.5 (1), b = 49.0 (1), c = 120.4 (1) Å, V = 2 861 000 $Å^3$, and the other belongs to the monoclinic space group *C*2, with unit-cell parameters a = 102.4 (1), b = 49.1 (1), c = 74.4 (1) Å, β = 129.0 (1)° and V = 290 000 $Å^3$. In both cases, the asymmetric unit contains a protein dimer and 35% solvent.

2.2. X-ray data collection

Low-temperature native data for the orthorhombic crystals were collected to 1.75 Å resolution at EMBL, c/o DESY, Hamburg on the X11 synchrotron beamline with an X-ray wavelength of 0.9057 Å using a MAR image-plate scanner. Data integration and scaling were performed using the *HKL*2000 system (Otwinowski & Minor, 1997). A total of 29 383 unique data were measured at 100 K from a crystal of dimensions 0.2 × 0.2 × 0.2 mm (R_{int} = 2.6%, completeness = 99.1%) mounted in a loop.

Apart from the data from the native orthorhombic crystal, all other diffraction data were collected at NSLS beamline X9B using an ADSC Quantum 4 CCD detector at a temperature of 100 K. These data were also processed with *HKL*2000. Various wavelengths were utilized as appropriate in order to obtain a meaningful anomalous scattering signal from different anomalous scatterers. The anomalously scattering halide ions were incorporated by soaking crystals in a cryosolution containing ∼1 *M* NaBr or KI for 30–60 s prior to freezing. Reflection data statistics are shown in Table 1.

2.3. Self-rotation function and non-crystallographic symmetry

The self-rotation function was calculated with the program *XPREP* (Sheldrick *et al.*, 2001). Fig. 1(*a*) shows a stereographic projection along the *z* axis of the $\kappa = 180^\circ$ section for the orthorhombic data, showing the presence of a twofold non-crystallographic axis bisecting the angle γ of the cell. Figs. 1(*b*) and 1(*c*) show stereographic projections along the *y* axis of the $\kappa = 70$ and $\kappa = 110^\circ$ sections for the monoclinic data. Four monomers describe a distorted fourfold non-crystallographic axis parallel to the crystallographic twofold axis.

2.4. Structure refinement

A standard refinement protocol was followed to obtain the final models used for mean phase-error estimation and map correlation-coefficient evaluation in the different phasing approaches. In both the monoclinic and orthorhombic cases, the model was refined isotropically against F^2 with *SHELXL*97 (Sheldrick & Schneider, 1997). 95% of the measured data to 1.75 Å were used for conjugate-gradient least-squares minimization, keeping a random set of 5% for cross validation with $R_{\rm free}$ (Brünger, 1992). In addition, local non-crystallographic symmetry restraints were applied to corresponding atoms of the two monomers for 1–4 distances and for isotropic displacement parameters (Usón *et al.*, 1999). Figures were drawn with *XTALVIEW* (McRee, 1999), *Raster3D* (Merritt & Bacon, 1997), *SHELXTL* (Sheldrick, 1990) and *SHELXPRO* (Sheldrick & Schneider, 1997). The phase quality attained in the different phasing approaches was evaluated through weighted mean-phase errors and map correlation coefficients (Lunin & Woolfson, 1993).

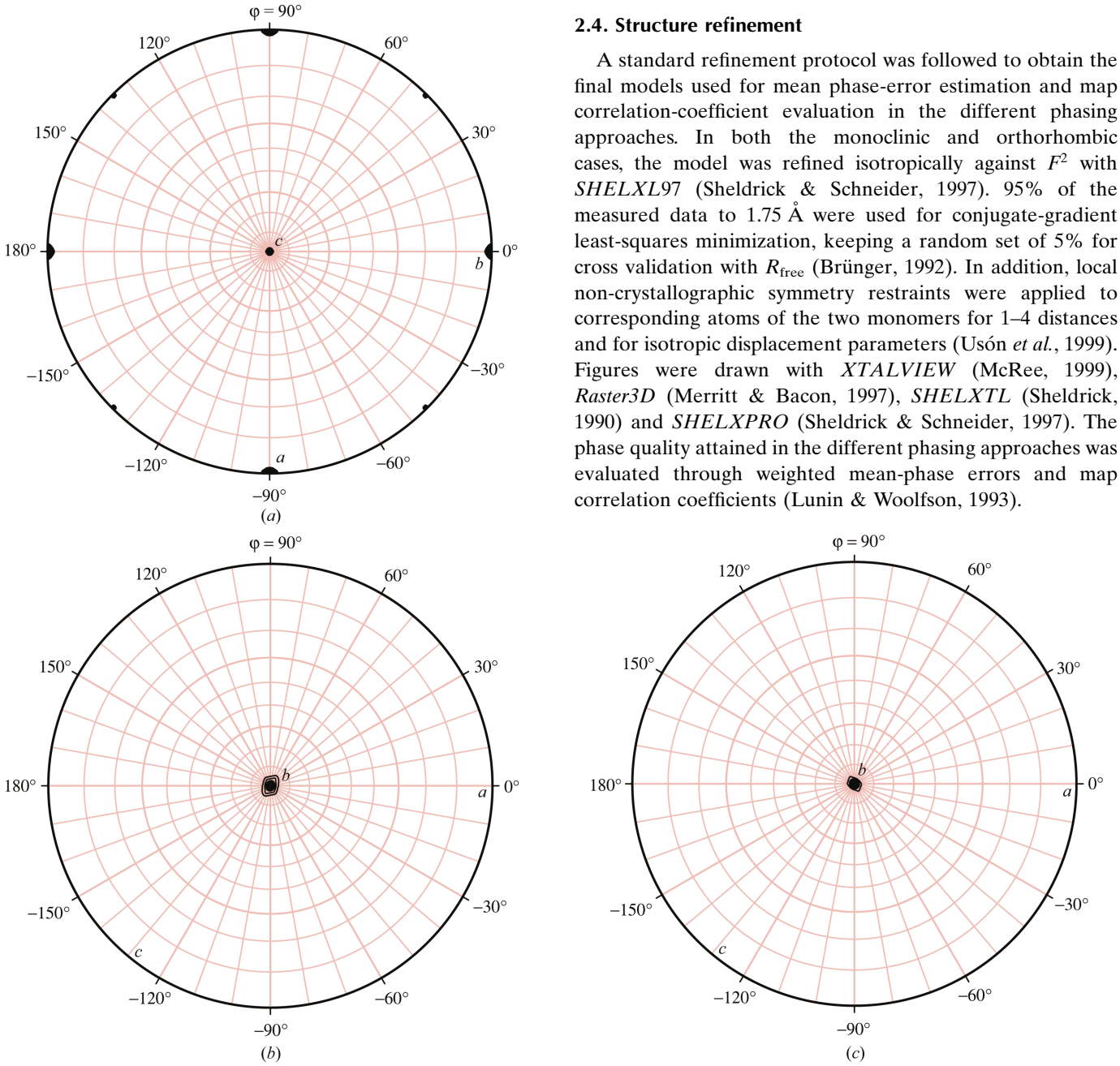


Figure 1
Stereographic projections (*a*) along the *z* axis of the $\kappa = 180^\circ$ section of the self-rotation function for the orthorhombic $P2_12_12_1$ crystal form, showing the presence of a non-crystallographic twofold axis in a direction bisecting the *x* and *y* axes; (*b*) along the *y* axis of the $\kappa = 70^\circ$ section of the self-rotation function for the monoclinic *C*2 crystal form; (*c*) along the *y* axis of the $\kappa = 110^\circ$ section of the self-rotation function for the monoclinic *C*2 crystal form. $\omega = 0^\circ$ at the centre of the projections and $\omega = 90^\circ$ at the circumference.

Table 2
Evaluation of anomalous signal.

(a) Statistics of the anomalous signal-to-noise ratio against resolution, $\langle|F^+ - F^-|/\sigma(F^+ - F^-)\rangle$

Resolution (Å)	∞–8	8–6	6–5	5–4	4–3.7	3.7–3.5	3.5–3.3	3.3–3.1	3.1–2.9	2.9–2.7	2.7–2.5	2.5–2.3	2.3–2.1
I soak	5.21	5.85	4.82	3.01	2.86	2.43	2.60	2.79	2.55	2.49	2.07	1.62	
Br soak, remote	2.69	2.80	2.21	1.64	1.50	1.50	1.40	1.48	1.41	1.26	1.24	1.17	1.14
Br soak, inflection	2.67	2.75	2.14	1.49	1.34	1.26	1.19	1.19	1.14	1.12	1.17		
Br soak, peak	3.62	3.36	2.74	1.91	1.52	1.41	1.36	1.38	1.25	1.21	1.27		
Sulfur 1	2.28	2.21	1.73	1.57	1.39	1.26	1.24	1.25	1.09	1.13	1.11	1.15	
Sulfur 2	1.58	1.65	1.36	1.23	1.07	1.13	1.13	1.09	1.22	1.19	1.15		
Sulfur merged	2.51	2.46	1.99	1.71	1.54	1.45	1.42	1.37	1.30	1.26	1.20	1.17	
Lysozyme	3.34	5.57	3.73	3.06	1.54	1.54	1.88	1.88	2.40	2.40	2.40	2.18	1.96
*C*2 Br peak	2.56	2.59	1.91	1.49	1.29	1.20	1.22	1.22	1.16	1.20	1.21	1.15	1.18
*C*2 Br inflection	1.81	1.81	1.43	1.29	1.31	1.31	1.37	1.36	1.33	1.33	1.37	1.34	1.24

(b) Correlation coefficient between the signed anomalous differences against resolution CC $\{(F - F^-)_i, (F^+ - F^-)_j\}$.

Resolution (Å)	∞–8	8–6	6–5	5–4	4–3.7	3.7–3.5	3.5–3.3	3.3–3.1	3.1–2.9	2.9–2.7	2.7–2.5	2.5–2.3	2.3–2.1
Br soak: remote/inflection	86.6	87.4	80.5	77.1	73.6	60.8	63.5	39.6	42.2	41.0	29.7	17.9	
Br soak: inflection/peak	91.8	90.3	81.8	80.7	78.6	62.0	67.6	57.8	48.5	39.2	38.2	22.6	
Sulfur 1/sulfur 2†	81.4	70.6	60.0	43.5	43.5	43.5	33.1	17.6	29.5	22.1	13.1	9.6	7.8
*C*2 Br inflection/peak	91.4	84.2	73.5	64.1	58.1	60.1	60.0	52.0	50.3	46.5	43.8	36.4	18.3

† Comparison of the anomalous differences from two different native crystals.

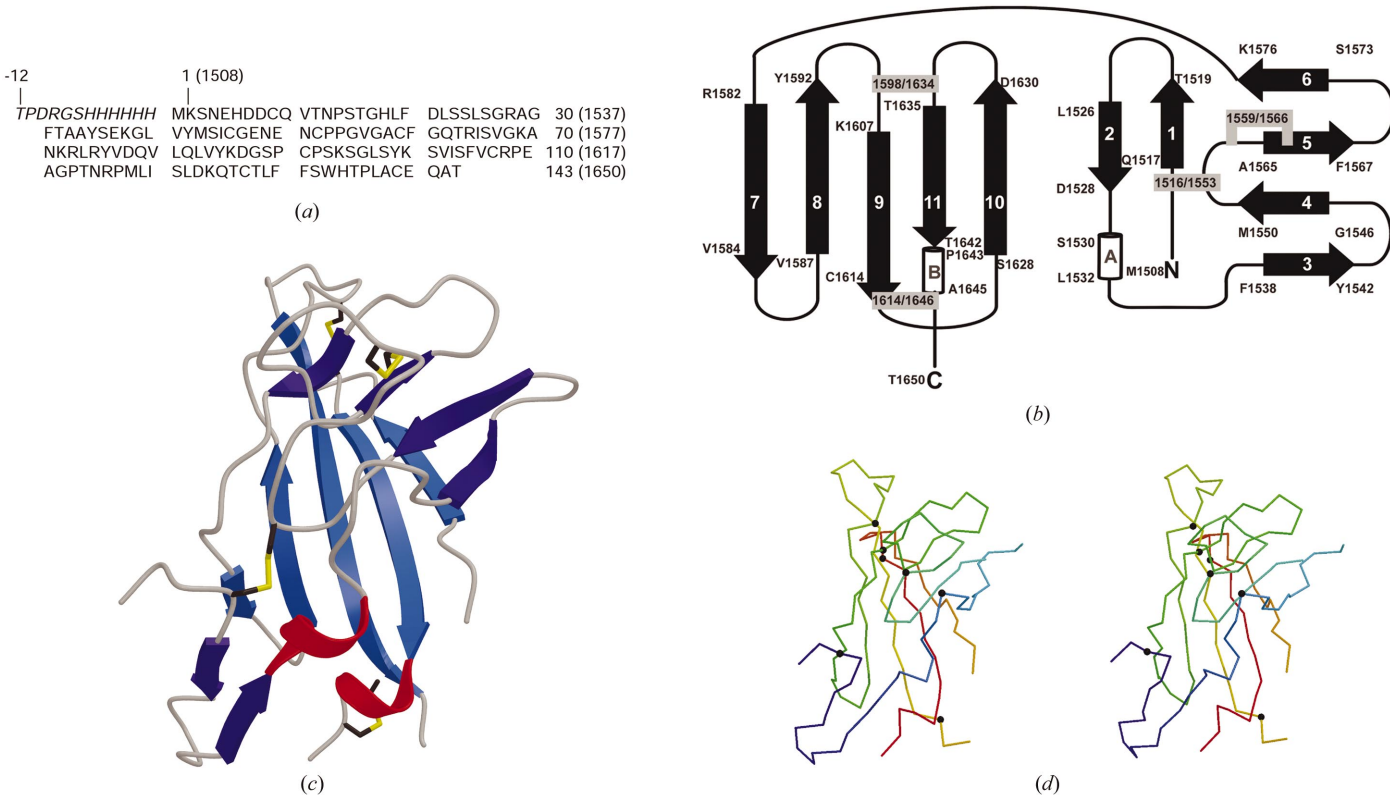


Figure 2
Structure of R11. (a) Amino-acid sequence of R11. The sequence starts N-terminally with the linker/His-tag sequence (italic) followed by repeat 11 of the human Man6P/IGFII receptor. The residue numbers of the complete Man6P/IGFII receptor are given in parentheses. (b) Topology of R11. β-strands are represented by black arrows and α-helices by cylinders. Disulfide bonds are enclosed by shaded dashed boxes. The numbering of residues is based on the sequence of R11. The β-strands are numbered from the N-terminus to the C-terminus; the α-helices are labelled A and B. The first and last residues are indicated in each secondary-structure element. (c) Ribbon diagram of the R11 structure, showing secondary-structural elements and disulfide bridges. (d) Stereo drawing of a C^α trace of R11, colour-ramped from the N-terminus (blue) to the C-terminus (red). Every 20th residue in the chain is marked.

3. Results and discussion

3.1. Data available

For the orthorhombic crystals, several data sets were collected to try different anomalous phasing approaches.

A native data set to 1.75 Å resolution was collected at a short wavelength (0.9057 Å) and shows no significant anomalous signal, as the protein contains only conventional amino acids.

Two complete data sets were measured on two native crystals in two different orientations, at a wavelength of 1.771 Å, aiming to maximize the inherent anomalous signal derived from the 22 S atoms present in the protein and from possible coordinated chloride anions arising from the crystallization buffer.

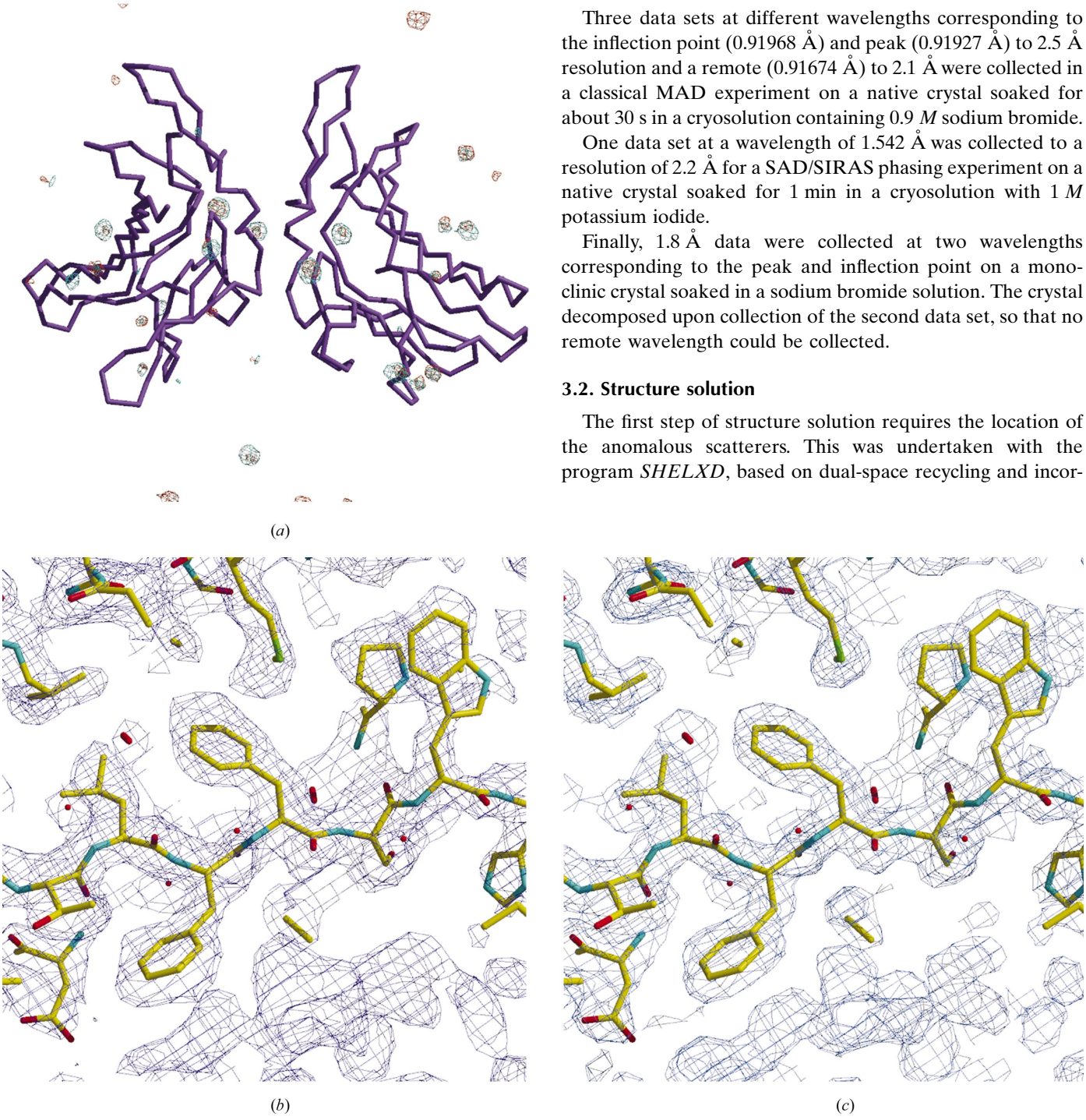
Three data sets at different wavelengths corresponding to the inflection point (0.91968 Å) and peak (0.91927 Å) to 2.5 Å resolution and a remote (0.91674 Å) to 2.1 Å were collected in a classical MAD experiment on a native crystal soaked for about 30 s in a cryosolution containing 0.9 *M* sodium bromide.

One data set at a wavelength of 1.542 Å was collected to a resolution of 2.2 Å for a SAD/SIRAS phasing experiment on a native crystal soaked for 1 min in a cryosolution with 1 *M* potassium iodide.

Finally, 1.8 Å data were collected at two wavelengths corresponding to the peak and inflection point on a monoclinic crystal soaked in a sodium bromide solution. The crystal decomposed upon collection of the second data set, so that no remote wavelength could be collected.

3.2. Structure solution

The first step of structure solution requires the location of the anomalous scatterers. This was undertaken with the program *SHELXD*, based on dual-space recycling and incor-



(*a*)

(*b*)

(*c*)

Figure 3
Experimental electron-density maps illustrating the anomalous scatterers in the structure of R11 with the final model superposed. (*a*) Fourier maps calculated with coefficients ΔF, $\varphi_{\rm calc} - 90^\circ$ for the whole molecule contoured at a 5σ level for the iodide SIRAS data (cyan) and for the averaged bromide SAD data (orange) in the $P2_12_12_1$ structure. (*b*) Weighted F_o map and final model after phasing with *SHARP* of the iodide SIRAS data. (*c*) Weighted F_o map and final model after *SHARP*/*DM* phasing of the iodide SIRAS data.

porating a combination of direct methods and Patterson methods (Usón & Sheldrick, 1999). The possibility of locating the anomalous scatterers depends on the significance of the anomalous signal present in the data. The maximum resolution that could be included was evaluated in the case of MAD data by determining the correlation between the signed anomalous differences ΔF at different wavelengths as a function of the resolution. For the native SAD data, two data sets collected on the same crystal were correlated. Otherwise, the signal-to-noise ratio in the anomalous signal present in a given data set can be estimated from the mean value of the ratio between the anomalous differences $|F^+ - F^-|$ and the estimated standard deviation of these differences, which can be taken from the processed data or modified taking into account the variance of all measured equivalents. Data were truncated at the resolution estimated to be the limit for significant anomalous signal, as higher resolution data add noise and tend to be overemphasized by direct methods because of the use of normalized structure factors. Table 2 summarizes the evaluation of the anomalous signal in the different experiments according to the above criteria and the resolution limit included in the calculations.

For the location of substructures of anomalous scatterers with *SHELXD*, only the internal loop, which relies on the strongest E magnitudes, is used. In this case, rather than starting with random atoms, as in *ab initio* structure solution, sets of atoms consistent with the Patterson solution were generated. For this purpose, 100 strong general peaks were selected from the sharpened Patterson as two-atom search fragments. The same random translation was applied to both atoms. For each of 99 999 random translations for the orthorhombic data and 9999 for the monoclinic data, the Patterson minimum function (PMF) was calculated for the two atoms and all their symmetry equivalents (Nordman, 1966). Before the first dual-space cycle, the number of starting atoms in the 'best' solution (highest PMF) was extended from two to the required N atoms by means of a full-symmetry Patterson superposition minimum function (PSMF; Buerger, 1959). Each grid point of the PSMF map was assigned the value of the PMF for all vectors between the two atoms and their symmetry equivalents and a dummy atom placed at that grid point, and the top N unique peaks of the PMSF were used as starting atoms for the dual-space recycling. From this point, alternating phase refinement by the tangent formula (Karle, 1968) with the selection of the $1.3N$ highest peaks in an E map and random elimination of $0.3N$ atoms was carried out for 20 cycles. Possible solutions were evaluated accordingly to the value of the PMF for the set of atoms (PATFOM), the correlation coefficient (CC; Fujinaga & Read, 1987) based on all data and a correlation coefficient based only on the weak data, meaning those that were left out of the dual-space recycling. This last 'weak' correlation coefficient is a sort of CC_{free}, as it is calculated on data that are independent of those used for substructure determination and therefore provides complementary information unless the weak data are too noisy, even though the approximation that ΔF corresponds to the heavy-atom amplitudes holds less well for weaker SAD or SIR data.

Once the substructure is solved, the atom coordinates can be used to calculate the raw experimental phases.

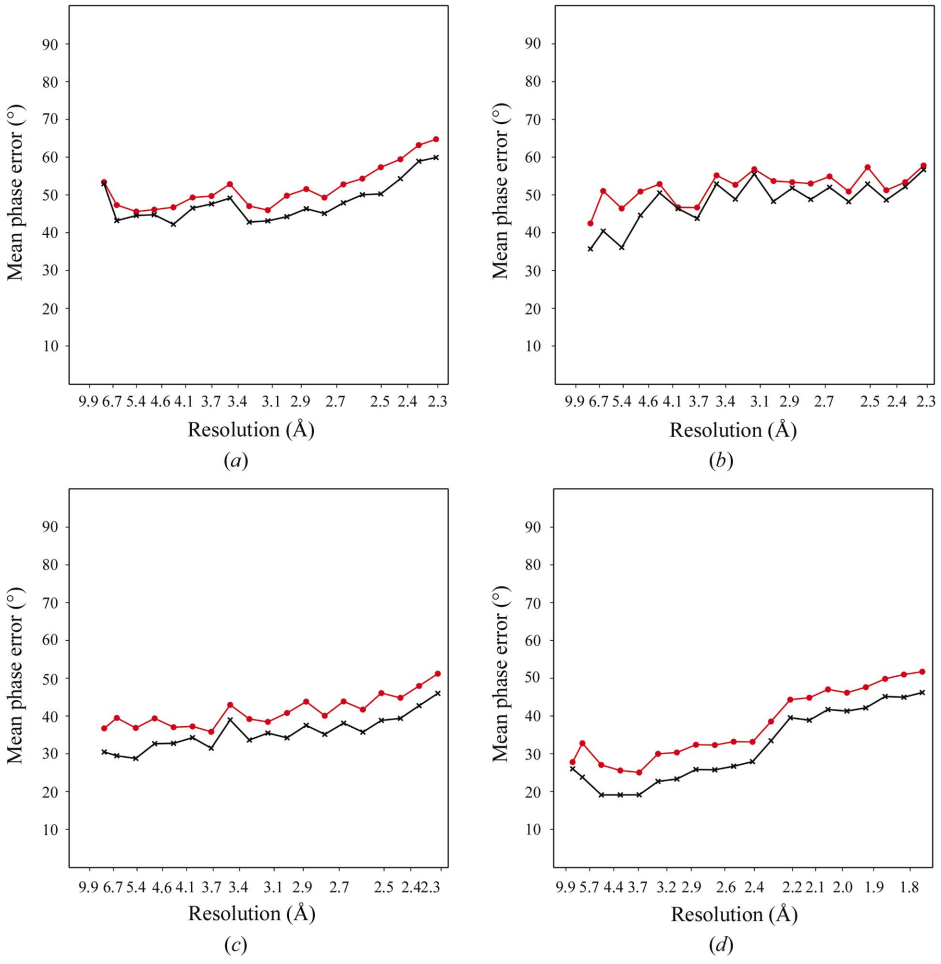


Figure 4
Plots showing the quality of the phases obtained in the different experiments, quantified as weighted mean-phase errors compared with the final refined structure. The weighted mean phase errors (wMPE) are plotted as red circles, with the F-weighted mean phase errors (FwMPE) plotted as black crosses. (*a*) Iodide SAD raw phases, (*b*) iodide SIR raw phases, (*c*) iodide SIRAS raw phases, (*d*) iodide SIRAS phases after *SHARP*/*DM*, (*e*) $P2_12_12_1$ bromide two-wavelength MAD (remote and inflection) raw phases, (*f*) $P2_12_12_1$ bromide SAD (with merged remote, peak and inflection data) raw phases, (*g*) $C2$ bromide SAD (peak) raw phases, (*h*) $C2$ bromide SAD (peak) phases after *SHARP*/*DM*.

3.3. Description of the structure

In both structures the asymmetric unit contains a dimer, whereas in both recently published crystal forms of this same protein (Brown *et al.*, 2002) only one monomer occupies the asymmetric unit. The average r.m.s. deviation between the C^{α} atoms of all 15 possible pairs of monomers is 0.58 Å, with the highest variability in the loop comprising residues 1598–1603. In the case of the orthorhombic structure, the two independent monomers are related by a non-crystallographic twofold axis perpendicular to z and bisecting the xy plane and a relative translation of 5 Å along the non-crystallographic axis. In the $C2$ structure, alternating 70 and 110° rotations relate independent monomers in a sort of distorted fourfold rotation axis parallel to the crystallographic b axis. The sequence, topology and three-dimensional structure of R11 are shown in Figs. 2(*a*), 2(*b*) and 2(*c*), respectively. A stereo drawing of the C^{α} trace is shown in Fig. 2(*d*). The first residue that can be determined in the electron density is His1501 in one of the monomers (Asp1502 in the remaining three). The preceding residues and the N-terminal tag and linker used for expression and purification cannot be detected in the electron density. The main-chain density is then continuous up to Glu1635 (Gln1636), with the exception of the loop containing residues Glu1605 (Ala1606) to Thr1609, which shows no interpretable electron density in any of the four monomers. In each monomer, the last two or three residues are missing. R11 consists almost entirely of β-strands, the only exception being two small segments of α-helix comprising residues 1518–1520 and 1631–1633. The β-strands form three β-sheets. Two β-strands are arranged in an antiparallel orientation preceding the first helix. The second sheet is composed of four antiparallel strands in an almost orthogonal orientation with respect to the third and largest sheet made up of five strands and located in the C-terminal half of the polypeptide chain, before the last short helix. The last two β-sheets form a kind of flattened β-barrel. The structure contains four disulfide bridges formed by cysteine residues conserved throughout all 15 repeats of the Man6P/IGFII receptor. Two of them are located in the N-terminal region and two in the C-terminal region of the protein. The first two link the loop comprising residues 1539–1552, located in the second β-sheet, with the first smaller β-sheet and with the third strand in the second β-sheet. The two disulfide bridges located in the C-terminal area link residues in the loops preceding and following strands 9 and 11 in the third β-sheet.

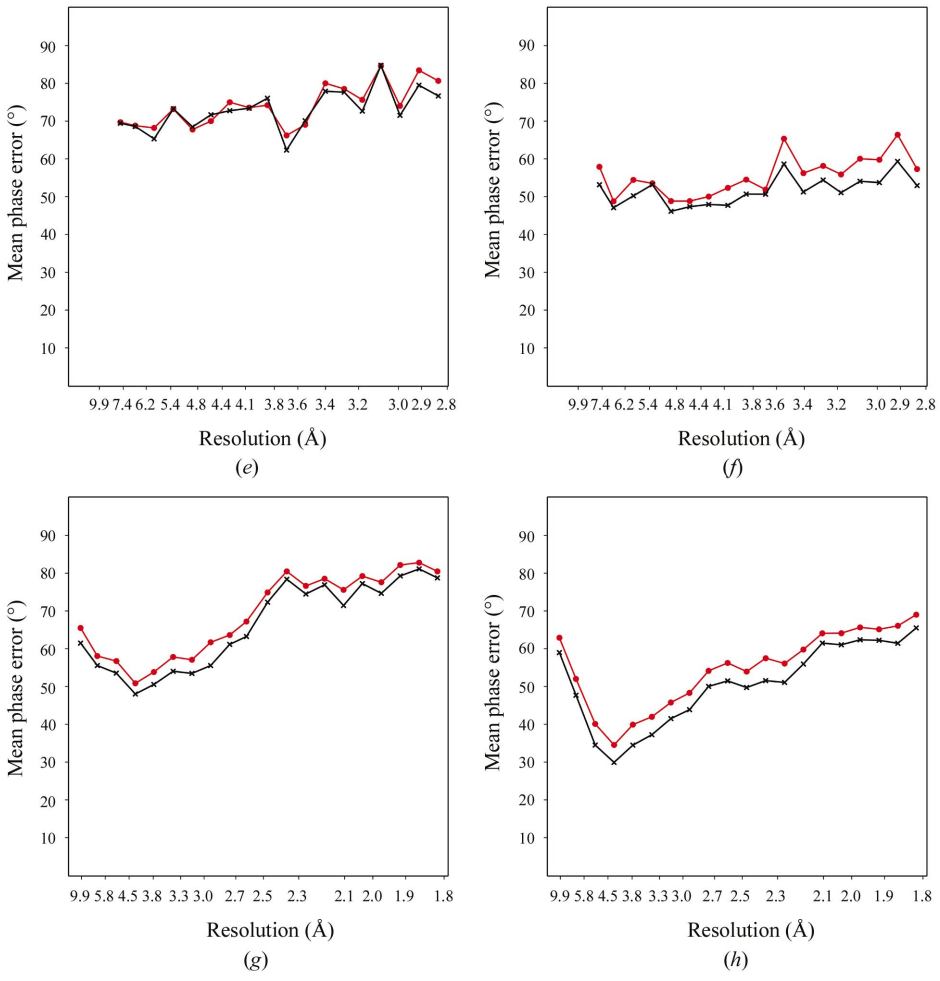


Figure 4 (continued)

3.4. Iodide derivative

The data from the iodide derivative showed significant anomalous signal throughout the entire resolution range to 2.3 Å (anomalous signal-to-noise ratio of 1.62 in the outer shell 2.5–2.3 Å; see Table 2).

Either SAD or SIR or SIRAS with the iodide and the native 1.75 Å data can be used to locate the positions of the I atoms, leading in this case to similar substructures, as the anomalous signal is significant and the isomorphism to the native data is acceptable throughout the resolution range.

Essentially correct solutions were reached in over 90% of the trials with either SAD, SIR or SIRAS data. All solutions showed eight strong iodide sites and several weaker sites. When the phases from at least 20 solutions were averaged (taking origin and enantiomorph shifts into account), a total of 16 sites could be identified (see Fig. 3*a*). The combined phases were used directly for phase error and map correlation-coefficient calculation. Alternatively, starting with only the eight strong sites, phase estimation with *SHARP* leads to location of the

remaining positions in the maximum-likelihood maps and eventually to the same solution. Experimental SIRAS maps after *SHARP* and *DM* are shown in Figs. 3(*b*) and 3(*c*), respectively.

Fig. 4 shows the weighted mean phase error (wMPE), *F*-weighted mean phase error (FwMPE) and map correlation coefficient (mapCC) values plotted against resolution for the phases obtained with the iodide derivative in the SAD, SIR and SIRAS experiments. SIRAS provided the best raw phases, with average wMPE = 41.7°, FwMPE = 34.9° and mapCC = 0.700 values for data to 2.3 Å resolution. SIR (wMPE = 52.5°, FwMPE = 47.7°, mapCC = 0.500) and SAD (wMPE = 61.6°, FwMPE = 55.6°, mapCC = 0.453) phases are somewhat poorer. This is as expected because the SIR and SAD phases were centroid phases, with no attempt to resolve the twofold ambiguity. This results suggest that in the case of the iodide-soaking experiment the isomorphous signal was somewhat stronger than the anomalous signal. In any case, phases can be further improved if they are estimated in the maximum-likelihood framework provided by *SHARP* (de La Fortelle & Bricogne, 1997) and subjected to density modification with *DM* (Cowtan, 1994) taking the non-crystallographic symmetry derived from the heavy-atom positions into account (wMPE = 38.9°, FwMPE = 27.4°, mapCC = 0.872 for the full resolution range of the native data to 1.753 Å). Fig. 4(*d*) shows the corresponding figures of merit after *SHARP*/*DM* treatment for the SIRAS case.

3.5. Bromide data

The anomalous signal in all three bromide data sets is significant and consistent to 2.8 Å (as shown by the correlation coefficient between signed anomalous differences for each pair of data sets). Therefore, it was unexpected that MAD F_A data derived from the three data sets did not lead to determination of the Br-atom positions. Neither did F_A values from a combination of the remote and peak data sets provide a solution for the substructure; only with the combination of the remote and inflection-wavelength data sets could three correct sites be determined. The raw phases derived from this last solution (wMPE = 74.4°, FwMPE = 72.5°, mapCC = 0.205) are still of very poor quality (see Fig. 4*e*).

It is therefore striking that combining all three data sets to calculate the anomalous ΔF differences for a SAD approach results in the easy location of 12 main sites and an additional weaker one, which leads to much better quality phases (wMPE = 55.5°, FwMPE = 51.6°, mapCC = 0.516) (see Fig. 4*f*).

These results were further confirmed by testing the quality of the phases derived from the 13 bromide sites located in the averaged SAD experiment and F_A and phase shifts between substructure and protein from the combined remote and inflection data sets (which provided the best result in the MAD phasing attempts). The figures of merit show that the maps were again unlikely to be interpretable (wMPE = 68.2°, FwMPE = 65.7°, mapCC = 0.315). This points to problems with the quality of the F_A values in the MAD experiment as the cause of the failure in the location of the substructure of anomalous scatterers. Indeed, dispersive differences could be more susceptible to inferior data quality, radiation damage and wavelength accuracy. Determination of the precise wavelength values for peak and inflection points might be affected by the presence of the free halides in the solvent region, affording different values from those corresponding to the weakly coordinated halides in the substructure. This suggests that in the case of weakly coordinated anomalous scatterers, as with halides in the protein solvent, it may be more straightforward to locate the anomalous substructure from SAD data with high redundancy than from MAD F_A data collected for the same total time.

The bromide data were in this case not sufficiently isomorphous to the native data for satisfactory SIR or SIRAS experiments, so these will not be discussed here further.

3.6. Phasing attempts on the anomalous signal of sulfur and chlorine alone

Phasing a protein through only the inherent anomalous signal derived from the S atoms present in cysteines and methionines and the additional signal of chloride in the ordered solvent region was possible for a lysozyme low-temperature synchrotron data set with average redundancy of 22 collected at a wavelength of 1.542 Å (Dauter *et al.*, 1999; in this case, even at 2.1 Å resolution the anomalous signal-to-noise ratio was 1.96 and it was significant to 1.7 Å; see Table 2).

In the data collected for our test case, despite using a long wavelength (1.771 Å) to maximize the anomalous signal, the anomalous signal-to-noise ratio as evaluated for the merged data set is significant only to 2.9 Å. Also, the correlation coefficient between both data sets collected on the same crystal shows values above 30% only between infinity and 3.3 Å (see Table 2).

The lower symmetry of our crystals ($P2_12_12_1$ compared with $P4_32_12$) and the single-circle goniometer also made it impracticable to collect data with such extremely high redundancy as in the lysozyme case.

Another decisive limitation is the resolution achieved, which in this case was limited to 2.3 Å. As S atoms in disulfide bridges are only about 2.05 Å apart, it should not be possible to resolve them, which is a requirement for success in methods derived from structure solution at atomic resolution.

Therefore, we started by testing on the anomalous lysozyme data set truncated to lower resolution the possibility of finding the location of the disulfide bridges as 'superatoms' with *SHELXD*. Indeed, with data truncated to resolutions between 2.5 and 3.2 Å it was possible to locate the centroids of the disulfide bridges, both methionine sulfurs and even a chloride anion in the solvent. At a resolution of 3.0 Å, these seven peaks are among the eight strongest selected in the solutions. At 3.2 Å, they are still present among the 14 strongest peaks, but discrimination between correct sites and spurious peaks becomes increasingly difficult. Also, the differences in figures of merit between solutions and non-solutions decreases with increasing resolution and it is the correlation coefficient rather

than the PATFOM figure of merit that characterizes correct solutions.

In the R11 case no sulfur substructure could be determined with data to 2.5 Å. The anomalous signal is significant but is presumably too weak to locate the S atoms or disulfide bridges, as an analysis of the anomalous maps (Kraut, 1968) calculated with coefficients ΔF, $\varphi_{\rm calc} - 90^\circ$ revealed. Indeed, only one of the disulfide bridges in each monomer showed density above the 5σ level, the remaining six disulfide bridges appearing at the 3σ level. Otherwise, the map appears to have a very low noise level, as the peaks corresponding to the disulfide bridges are indeed the highest peaks. On the other hand, it is well known that phases are more important than magnitudes in the calculation of a map and as the phases are derived from the final model of the refined structure, conclusions on the quality of the anomalous signal in the original data based on such a map may well be deceiving. Additionally, the final model completely lacks one of the three methionines in each monomer, located close to the N-terminus for which the electron density is missing; the remaining two are disordered. This also suggests that in the case of this structure, the selenomethionine approach initially but unsuccessfully attempted could never have succeeded!

Interestingly, in the case of the lysozyme data truncated to 2.5 Å resolution, the structure could be phased with *SHARP* starting only from the positions of the centroids of the disulfide bridges present. The remaining S and Cl atoms can be located in the difference maps and the actual S-atom sites in each disulfide bridge split from the centroids upon examination of the negative and positive maxima in the maximum-likelihood maps or upon anisotropic refinement of the centroid sites. The electron-density map thus obtained can be traced directly or further improved through density modification.

3.7. Bromide derivative of the *C*2 crystal form

For the monoclinic crystal form, only two data sets were recorded from a crystal soaked with sodium bromide. Radiation damage is apparent in the second data set corresponding to the inflection point. Nevertheless, the same strategy outlined in the case of the orthorhombic data sets leads to determination of a substructure composed of 12 bromide sites with the SAD values derived from the peak data set. These render raw phases of wMPE = 65.0°, FwMPE = 58.4° and mapCC = 0.350 (see Fig. 4*g*) and leads after the use of *SHARP*/*DM* to an interpretable map characterized by figures of merit of wMPE = 56.1°, FwMPE = 47.6° and mapCC = 0.638 (see Fig. 4*h*).

4. Conclusions

In all cases, if the anomalous substructure could be determined then the rest of the phasing process was successful, including the case in which centroids of the disulfide bridges rather than individual S atoms are determined.

As far as our test case of a previously unknown protein with 310 amino acids in the asymmetric unit is concerned, the very weak anomalous signal from sulfur was insufficient to solve the structure. However, some anomalous signal was found, so further improvement in data-collection or computational methods may make this approach feasible in the future.

Phasing after a short soak with a buffer containing a halide salt is in any case easier and more likely to succeed. The limit to the number of partially occupied iodide sites *SHELXD* could find for larger structures remains to be explored. Both halides have been shown to occupy the same sites in this structure. Data quality is decisive for successful location of the anomalous substructure, so if data-collection time or crystal stability are limiting factors, it might be better to optimize redundancy for a data set with good anomalous signal than to aim for several data sets in a MAD experiment.

Iodine is a better choice for isomorphous replacement purposes and has a considerable anomalous signal, especially if a long wavelength can be selected, but is not suitable for MAD.

We are grateful to Clemens Vonrhein and Gerard Bricogne for support with *SHARP*. We thank the Deutsche Forschungsgemeischaft (DFG) and the Fonds der chemischen Industrie for financial support. Data collection at the synchrotron beamlines X9B at NSLS and X11 in Hamburg at the EMBL Outstation c/o DESY with assistance from Dr Alexander Popov during data collection is gratefully acknowledged; synchrotron access at DESY was supported by the European Union grant (QLRI-CT-2000-00398).

References

Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: Wiley.
Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W.-R., Furey, W. F., Sax, M. & Wang, B. C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240–4244.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. & Derewenda, Z. (2000). *Structure*, **8**, 1137–1146.
Fujinaga, M. & Read, R. (1987). *J. Appl. Cryst.* **20**, 517–521.
Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
Grimme, S., Höning, F., von Figura, K. & Schmidt, B. (2000). *J. Cell Biol.* **275**, 33697–33703.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Karle, J. (1968). *Acta Cryst.* B**24**, 182–186.
Kraut, J. (1968). *J. Mol. Biol.* **35**, 511–512.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
McRee, D. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Nordman, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29–38.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sheldrick, G. M. (1990). *SHELXTL Software Package for the Determination of Crystal Structures, Release* 6.12. Bruker Analytical X-ray Systems, Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sheldrick, G. M. & Schneider, T. R. (2001). *Methods in Macromolecular Crystallography*, edited by D. Turk & L. Johnson, pp. 72–81. Amsterdam: IOS Press.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2001). *International Tables for Macromolecular Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, ch. 16, pp. 333–345. Dordrecht: Kluwer Academic Publishers.
Smith, J. L. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 211–225. Dordrecht: Kluwer Academic Publishers.
Usón, I., Pohl, E., Schneider, T. R., Dauter, Z., Schmidt, A., Fritz, H.-J. & Sheldrick, G. M. (1999). *Acta Cryst.* D**55**, 1158–1167.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–111.